High-throughput cloning and expression in recalcitrant bacteria

Geertsma, Eric R.; Poolman, Bert

Published in:
Nature Methods

DOI:
10.1038/NMETH1073

Publication date:
2007

High-throughput cloning and expression in recalcitrant bacteria

Eric R Geertsma & Bert Poolman

We developed a generic method for high-throughput cloning in bacteria that are less amenable to conventional DNA manipulations. The method involves ligation-independent cloning in an intermediary *Escherichia coli* vector, which is rapidly converted via vector-backbone exchange (VBEx) into an organism-specific plasmid ready for high-efficiency transformation. We demonstrated VBEx proof of principle for *Lactococcus lactis*, but the method can be adapted to all organisms for which plasmids are available.

The heterologous expression of large multidomain assemblies and membrane proteins in a functional state is often problematic in the well-established expression hosts *E. coli*, and yeasts such as *Pichia pastoris* and *Saccharomyces cerevisiae*[1–4]. Alternative expression systems are urgently needed to overcome this major hurdle in structural genomics projects. Frequently, efficient DNA manipulations are the bottleneck for exploring the protein expression potential of new hosts, thereby preventing rapid and routine screening. Shuttle vectors that replicate both in *E. coli* and the alternative host offer a way out, but their use is complicated by the requirement for multiple origins of replication and selection markers. This increases the plasmid size and often compromises stable propagation. *L. lactis* is widely recognized as an attractive alternative to expression systems based on *E. coli* and yeasts[5,6]. But low efficiency of gene manipulations in the organism and the instability of *L. lactis*–*E. coli* shuttle vectors have seriously hampered the expression screening in *L. lactis*. Additionally, the lack of an efficient cloning procedure has restricted the preparation of large gene libraries for directed evolution studies in *L. lactis*; similar limitations have prohibited the full exploitation of other microorganisms as expression and screening hosts.

We now present a generic cloning strategy, compatible with high-throughput manipulations, which generates a native plasmid vector optimal for the expression host and devoid of alien (for example, *E. coli*–derived) elements. The VBEx procedure presented here is specific for cloning in *L. lactis*, but may be readily adapted to all organisms for which a plasmid, selection marker and transformation method are available. Using VBEx, we generated over 300 gene constructs and assessed protein expression levels at a rate of 48 constructs per week without robotics.

To facilitate the initial steps in cloning numerous open reading frames, we used a ligation-independent cloning (LIC) procedure[7]. Contrary to methods that rely on recombination events (for example, Gateway[8] or the Univector Plasmid-fusion System (UPS)[9]), ligation-independent cloning is less restricted in the design of the sequences flanking the gene(s). Therefore, the cloning-related sequences attached to the recombinant protein(s) can be minimized. The cloning procedure involves linearization of the vector by restriction at a unique *Swa*I site in the middle of the LIC cassette and exploits a feature of T4 DNA polymerase: with only one nucleotide present, T4 DNA polymerase exhibits 3′ to 5′ exonuclease activity until the point in the sequence where the available nucleotide can be incorporated. As the 3′ sequences adjacent to the *Swa*I site are devoid of dCMP, T4 DNA polymerase treatment in the presence of dCTP will yield long, defined single-stranded overhangs. Complementary single-stranded overhangs are created in a similar way in the PCR product to be cloned (**Fig. 1a** and **Supplementary Note** online). The LIC cassettes are preceded by promoter regions specific for the intended expression host. Here we used the P_{BAD} and P_{NIS} promoters of the *E. coli* and *L. lactis* expression plasmids pBAD24 (ref. 10) and pNZ8048 (ref. 11), respectively. Ligation-independent cloning of PCR products proved highly efficient for the *E. coli* vectors, indicating that the length and composition of the complementary overhangs of vector and insert suffice for formation of stable heteroduplexes. In contrast, direct transformation of *L. lactis* with the stable heteroduplex of the pNZ8048-derived LIC vector pNZxLIC and a compatible insert yielded no or only very few positive clones (data not shown).

To overcome the poor cloning efficiencies in *L. lactis*, we devised a new strategy, VBEx, which allowed the initial (high-throughput) cloning to be done in *E. coli*, but avoided the use of shuttle vectors. Although we have used VBEx in combination with LIC, the strategy is fully compatible with Gateway[8], UPS[9] and variations on LIC (for example, enzyme-free cloning[12] and sequence and ligation–independent cloning (SLIC)[13]). The VBEx method relies on the bisection of a *bona fide* plasmid vector of the expression host into two parts, thereby separating the selection marker and the origin of replication. For *L. lactis*, we separated the origin of replication of the pNZxLIC plasmid (pSH71 replicon) from the chloramphenicol resistance gene (*cat*; **Fig. 1b**) by constructing two new plasmids, each with one of the segments. We fused the other segment containing the *cat* and LIC sequence to the backbone of a vector containing an *E. coli* origin of replication and the β-lactamase resistance gene (*bla*). The resulting plasmid, pRExLIC, allows the LIC manipulation to take place in *E. coli*. We fused the other segment containing the *L. lactis* origin of replication to an erythromycin selection marker (yielding plasmid pERL), which allows replication and selection in the expression host.

Department of Biochemistry, Groningen Biomolecular Sciences and Biotechnology Institute, and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG, Groningen, The Netherlands. Correspondence should be addressed to B.P. (b.poolman@rug.nl).

RECEIVED 4 MAY; ACCEPTED 27 JUNE; PUBLISHED ONLINE 22 JULY 2007; DOI:10.1038/NMETH1073

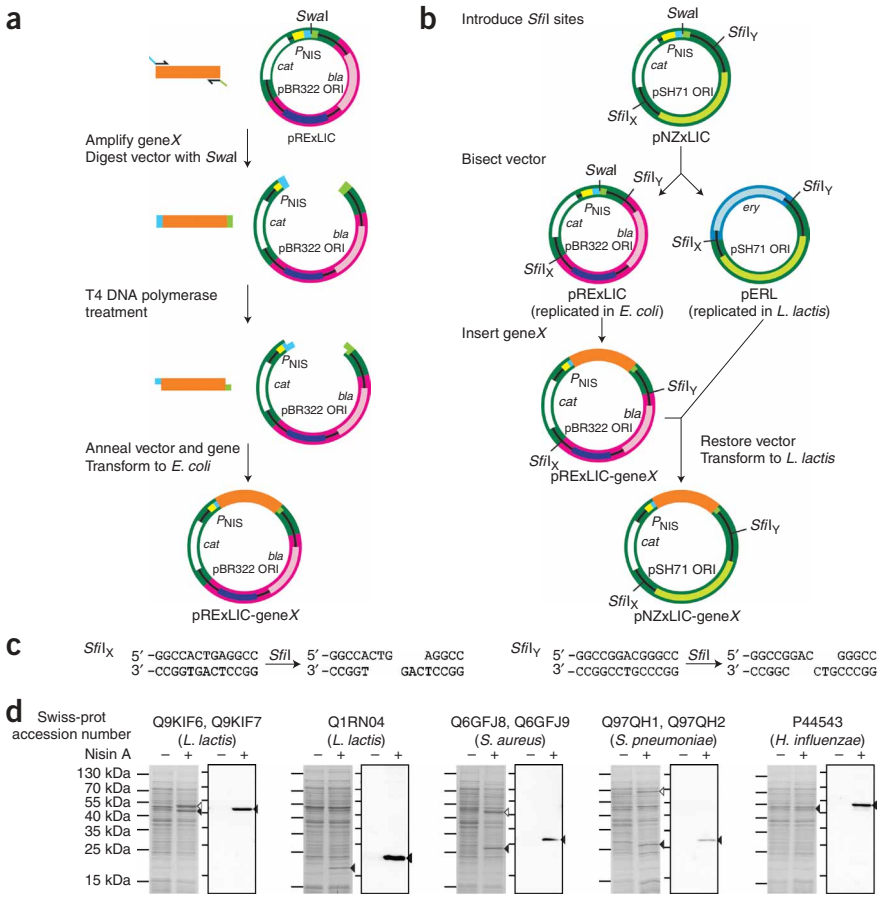

Figure 1 | High-throughput cloning in recalcitrant bacteria using LIC and VBEx. (**a**) In the LIC procedure, gene *X* is amplified using primers containing LIC-specific overhangs (blue and green). The plasmid is linearized by *Swa*I restriction in the LIC cassette. Single-stranded overhangs of the PCR product and vector are generated using T4 DNA polymerase (see **Supplementary Note**). The complementary overhangs of PCR product and vector anneal upon mixing. The resulting heteroduplex is transformed efficiently to *E. coli*. (**b**) In the VBEx strategy, the *L. lactis* expression vector pNZxLIC is cut at the two introduced *Sfi*I sites. Plasmid pERL consists of the pSH71 replicon (light green) from pNZxLIC fused to an erythromycin marker (blue). Plasmid pRExLIC consists of the *cat* marker (white) and LIC sequence from pNZxLIC, fused to the *E. coli* pBR322 replicon (purple) and the *bla* marker (pink). This vector is subjected to the LIC procedure (**a**); then the pNZxLIC vector is restored by mixing pERL and pRExLIC–gene *X*, digestion with *Sfi*I, ligation and selection on the ability to replicate in *L. lactis* in the presence of chloramphenicol. (**c**) The *Sfi*I sites flanking both segments of the vectors yield different, nonpalindromic overhangs. (**d**) Overexpression of membrane proteins in *L. lactis*. Membrane vesicles were analyzed by SDS-PAGE (left image in each set) or immunoblotted with an antibody to the His tag (right). Filled arrows indicate His-tagged and open arrows non-tagged subunits or proteins.

Rapid and high-throughput compatible reconstruction of the original *L. lactis* expression plasmid from the relevant segments of pRExLIC and pERL is assured by flanking the ends of each segment with distinct *Sfi*I restriction sites ($SfiI_X$ and $SfiI_Y$; **Fig. 1b**). DNA cleavage by *Sfi*I generates a 3′ overhang that can be composed of any combination of three nucleotides[14]. The two *Sfi*I sites used yield different, nonpalindromic overhangs that are not compatible with each other (**Fig. 1c**). The combination of (i) both halves of the original expression vector having properties that can be selected for, and (ii) different 3′ overhangs after *Sfi*I digestion at either side of each plasmid segment, ensures minimal sample handling and allows automation of the method. As each half of the expression vector has unique selectable properties, no gel electrophoresis and purification of the *Sfi*I-fragments is required. Selection for the ability of the plasmid to replicate in *L. lactis* in the presence of chloramphenicol permits unique recovery of only the original expression vector from a complex mixture of *Sfi*I-digested pRExLIC and pERL plasmids. Furthermore, no change of buffers between the digestion and ligation reaction is needed. Mere supplementation of the *Sfi*I-buffer with ATP suffices for the T4 DNA ligase to be fully active.

In practice, the entire procedure can take place in a single tube in 3 h. After joined *Sfi*I digestion of pRExLIC and pERL (80 min at 50 °C), and subsequent heat inactivation of the restriction enzyme (20 min at 80 °C), ligation (60 min at 20 °C) of the fragments is initiated by the addition of ATP and T4 DNA ligase. Upon thermal inactivation of the T4 DNA ligase (20 min at 65 °C), aliquots of the mixture can be used for (electro-)transformation without further purification (**Fig. 1b**; a detailed protocol is available in **Supplementary Methods** online). For *L. lactis*, we routinely observe high, reproducible transformation efficiencies (~10^6 c.f.u./μg of DNA), in comparison to the low, variable efficiencies

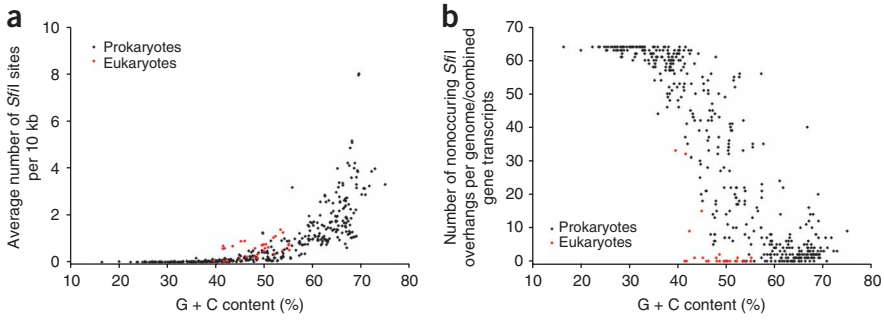

Figure 2 | Characteristics of *Sfi*I sites in prokaryotic and eukaryotic genomes. (**a**) The occurrence of *Sfi*I sites in 492 prokaryotic genomes and the combined gene transcripts of 30 eukaryotes as a function of the G+C content of the DNA. (**b**) The number of *Sfi*I overhangs that are not present in a genome or combined gene transcript and therefore available for VBEx as a function of the G+C content of the DNA (maximal 64). The full dataset is available in **Supplementary Data**.

of traditional restriction-ligation cloning ($\sim 10^2$ c.f.u./μg of DNA; unpublished observations and ref. 15). Notably, with the VBEx procedure, all *L. lactis* transformants obtained carry the gene of interest inserted in pNZxLIC. The overexpression of a small subset of membrane proteins in *L. lactis* is presented in **Figure 1d** (for details on the expression system, see refs. 5 and 11).

Notably, DNA sequences from virtually all sequenced genomes are compatible with the cloning strategy, because *Sfi*I sites (5′-GGCCNNNN^NGGCC-3′ (ref. 14)) are rare (**Fig. 2a** and **Supplementary Fig. 1** online). Analysis of all predicted open reading frames and gene transcripts of 492 archeal and bacterial chromosomes and 30 eukaryotic genomes, present in the National Center for Biotechnology Information (NCBI) databank (February 2007) and Ensembl (release 43), respectively, indicated that well over 92% of these transcripts do not contain any *Sfi*I sites (**Supplementary Table 1** online). Moreover, use of the method is not limited to transcripts free of *Sfi*I sites. As 64 different 3′ overhangs may be generated after *Sfi*I digestion, internal *Sfi*I sites with 3′ overhangs not matching those of the vector will result in a three-way or more ligation, but not form a bottleneck in the procedure. If needed, the vector can be readily adapted to use nonoccurring or extremely rare overhangs. In 89% of the genomes analyzed, at least two types of *Sfi*I overhangs did not occur (**Fig. 2b**). The remaining 11% of the genomes contain several types of low-occurrence *Sfi*I overhangs (**Supplementary Data** online).

To ensure the generality of the presented strategy toward inserts of different size, we compared *Sfi*I digestion rates of pRExLIC derivatives holding inserts up to 3.7 kb (data not shown) and observed complete digestion after 80 min of incubation. Furthermore, we demonstrated the absence of expression from the P_{NIS} promoter in the cloning host *E. coli*, using a sensitive activity-based assay (data not shown). The success of assembly and stability in the pRExLIC and pNZxLIC vectors of recombinant DNA from bacterial, plant and mammalian origin proved very high, and gene rearrangements have so far not been observed ($n > 300$).

In summary, we developed LIC-VBEx, a high throughput–compatible cloning system for *L. lactis* that has unprecedented high efficiency and can be readily adapted for use with any other expression host. Problems arising from shuttle vectors are avoided by using genuine expression plasmids. The developed LIC cassettes allow the tagging of the protein of interest with a cleavable 10His-tag at either the N or C terminus (vectors holding these cassettes and derivatives with alternative tags or fusion-partners are shown in **Supplementary Table 2** online). This procedure has been used in our laboratory to prepare over 300 gene constructs with high efficiencies (~90% for LIC, 100% for VBEx; data not shown). In a nonautomated setting, the full procedure from cloning to expression screening in *E. coli* and *L. lactis* took place at a rate of approximately 48 constructs per week.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We acknowledge D.J. Slotboom for critical reading of the manuscript, G.K. Schuurman-Wolters, C. Mulligan and D.J.S. for independent testing of the LIC-VBEx method, and the EU-FP6 programme (E-MeP; 504601) for funding.

AUTHOR CONTRIBUTIONS

E.R.G. designed and performed the experiments and analyzed the data; B.P. supervised the project; E.R.G. and B.P. wrote the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Grisshammer, R. & Tate, C.G. *Q. Rev. Biophys.* **28**, 315–422 (1995).
2. Tate, C.G. *et al. Biochim. Biophys. Acta* **1610**, 141–153 (2003).
3. Ding, H.T. *et al. Acta Crystallogr. D Biol. Crystallogr.* **58**, 2102–2108 (2002).
4. Lundstrom, K. *et al. J. Struct. Funct. Genomics* **7**, 77–91 (2006).
5. Kunji, E.R., Slotboom, D.J. & Poolman, B. *Biochim. Biophys. Acta* **1610**, 97–108 (2003).
6. Quick, M. & Javitch, J.A. *Proc. Natl. Acad. Sci. USA* **104**, 3603–3608 (2007).
7. Aslanidis, C. & de Jong, P.J. *Nucleic Acids Res.* **18**, 6069–6074 (1990).
8. Walhout, A.J. *et al. Methods Enzymol.* **328**, 575–592 (2000).
9. Liu, Q. *et al. Curr. Biol.* **8**, 1300–1309 (1998).
10. Guzman, L.M. *et al. J. Bacteriol.* **177**, 4121–4130 (1995).
11. de Ruyter, P.G., Kuipers, O.P. & de Vos, W.M. *Appl. Environ. Microbiol.* **62**, 3662–3667 (1996).
12. de Jong, R.N., Daniels, M.A., Kaptein, R. & Folkers, G.E. *J. Struct. Funct. Genomics*; published online **13** February 2007.
13. Li, M.Z. & Elledge, S.J. *Nat. Methods* **4**, 251–256 (2007).
14. Qiang, B.Q. & Schildkraut, I. *Nucleic Acids Res.* **12**, 4507–4516 (1984).
15. Surade, S. *et al. Protein Sci.* **15**, 2178–2189 (2006).